

天然有機化合物の NTG-活性辞書ツールの開発

- 構造類似性検索の機能強化 -

Making an NTG-activity dictionary of bioactive natural organic compounds

- An enhanced structural similarity search -

桂樹 哲雄^{*1}

Tetsuo Katsuragi

高橋 由雅^{*1}

Yoshimasa Takahashi

^{*1} 豊橋技術科学大学大学院工学研究科 情報・知能工学専攻

Department of Computer Science and Engineering, Toyohashi University of Technology

In the previous work, we developed an NTG-activity dictionary of bioactive natural organic compounds, which was aimed for providing a knowledge base for the topological molecular frameworks and the related biological activities. An NTG (Non-Terminal Vertex) is defined as a vertex graph which does not have any terminal vertex and any isolated vertex, and used to represent the skeleton structure of molecules which may be related to its biological activity. In this study, we added a new search feature using a prior search on the basis of the molecule's NTGs. The details of the NTG dictionary and the new search feature will be discussed with some examples.

1. はじめに

一般に、似た構造をもつ化合物は同一あるいは密接に関連する生理活性を発現することが知られている。先に、当研究室では、MDDR (MDL Drug Data Report) の薬化合物のデータを対象として、化合物の環構造に着目した Non-terminal Vertex Graph (NTG) [Takahashi 04] に基づいて構造と生理活性の相関性を整理し、検索することができる知識ベースを開発した [能登 08]。さらに、その後対象を天然有機物として、KNAPSAcK Core データベース [Afendi 12] の構造情報と KNAPSAcK Activity データベース [Nakamura 14] の生理活性データを用いた同様の知識ベースを開発した [Katsuragi 15]。これらの知識ベースでは、事前の部分構造定義を必要としない記述子として TFS (Topological Fragment Spectra) [Takahashi 98] を用いて、構造の類似性を計算した。しかしながら、この方法は、同一の骨格構造を有する化合物群に対して、これを優先的に反映した類似性評価を与えるとは限らない。

そこで、本研究では、当研究室で別途開発した、NTG を用いて分子の骨格構造を優先的に評価する類似性評価手法 [和田 05] を、上述の天然有機物の構造-生理活性知識ベース上に実装した。

2. 分子骨格と特徴記述子

2.1 Non-terminal Vertex Graph (NTG)

分子骨格を抽出するために、本研究では Non-terminal Vertex Graph (NTG) [Takahashi 04] を用いる。NTG は、分子の環構造に着目した部分グラフで、次数が 1 以下の頂点原子を持たないグラフと定義される。NTG は、頂点原子と結合タイプの種類に重みづけをするか否かで、以下の 4 つの表現レベルで定義することができる (図 1)。

- NTG / SG: 単純グラフ
- NTG / VG: 頂点原子の種類の重みつきグラフ
- NTG / EG: 結合辺の種類の重みつきグラフ

連絡先: 桂樹 哲雄, 〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1 豊橋技術科学大学大学院 情報・知能工学系, katsuragi@cs.tut.ac.jp

- NTG / CG: 頂点原子, 結合辺の種類の重みつきグラフ
さらに、厳密には NTG ではないが、以下を追加で定義する。
- NTG / DG: NTG / CG に隣接する 2 重結合を加えたグラフ

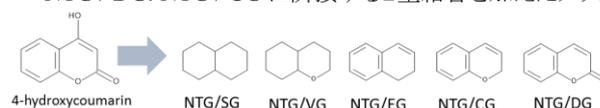


図 1 NTG の例 (4-hydroxycoumarin)

2.2 Topological Fragment Spectra (TFS)

Topological Fragment Spectra (TFS) [Takahashi 98] は、構造情報の特徴づけるための数値的な記述手法の一つである。本研究では、TFS を用いて化合物の構造類似性を評価する。TFS の生成方法は以下のとおり。対象の構造について、可能なすべての部分構造を列挙し、それぞれの部分に対して数値的特徴づけを行う。この数値の出現頻度のヒストグラムが TFS であり、これによって構造情報を多次元のパターンベクトルとして扱うことができる。部分構造の特徴づけの方法を変えることで、さまざまな特徴量 (各頂点の次数和, 質量数など) を扱うことができる (図 2)。

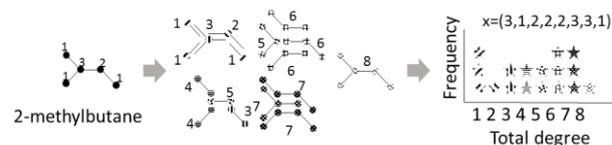


図 2 各頂点の次数和を特徴量とした場合の TFS の例 (2-methylbutane)

3. 分子の骨格構造を優先した類似性の評価手法

化合物の化学構造の類似性を評価する際に、分子の骨格構造を考慮するために、全グラフ (分子グラフそのもの) から生成した TFS の類似性と、抽出した NTG から生成した TFS の類似性の両面から評価を行う (図 3)。先行研究 [和田 05] の方法に従い、以下の通りに類似化合物の順位付けを行う。1) まず、抽出した NTG 間の類似性を評価し、NTG の類似性が高いものから順に順位をつける。2) 同型の NTG を持つ分子グラフが複数ある場合には、全グラフ間で類似度が高いものから順に順位をつける。

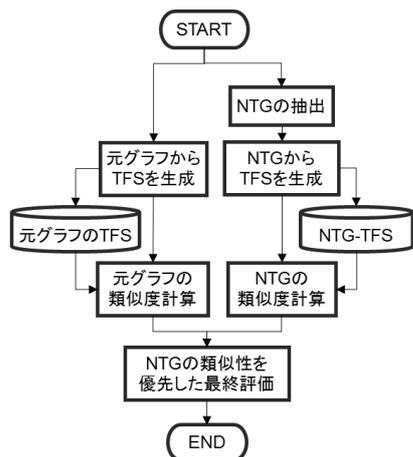


図3 骨格類似性を優先した構造類似性の評価方法

4. NTG-活性辞書ツールへの実装

4.1 天然有機物の NTG-生理活性知識ベース

天然有機物の NTG-生理活性知識ベース [Katsuragi 15] は、天然有機物を対象として、NTG とそれに関連する活性を紐づけた辞書ツールである。主な機能として、1) 入力化学構造から NTG を抽出し、その NTG に関連する生理活性を検索する、2) 入力化学構造から NTG を抽出し、登録された NTG との構造類似性検索を行い、構造が類似する NTG の持つ生理活性を列挙する、3) 生理活性から、その生理活性に関連する NTG を検索する、などの機能がある。

4.2 検索機能の強化

(1) 生理活性から関連する化合物の元構造を検索

先行研究では、NTG のみを対象とした構造検索機能が実装されていた。そこで、生理活性からその生理活性を持つ化合物の元構造を検索する機能を追加した。

(2) 類似性検索機能の対象に元構造を追加

先行研究では、類似性検索の対象が NTG のみであったため、登録されている化合物の元構造との類似度を検索できるように、上で示した骨格類似性を考慮に入れた評価手法による類似構造検索機能を追加した。

4.3 実データベースに対する類似性検索結果

クマリンは、植物の芳香成分の一つである。以下では、クマリンを入力構造として、機能強化した天然有機物の NTG-生理活性知識ベースを用いて骨格類似性を優先した構造類似性検索を行った例を示す(図4)。NTG の抽出には NTG / DG を用い、TFS の特徴量として、各頂点の質量数とを、類似度の計算方法として、Tanimoto 係数を用いた。図5は、検索結果の上位 1000 件について、クマリンとそれぞれの化合物における元構造間、NTG 間の類似性と、順位との関係を示す。NTG の類似度は、順位が下がるにつれて単調減少している。NTG の類似度は、1 番目から 226 番目までは 1.0 であり、クマリンが持つ NTG と同型の NTG を持つ化合物が、データベース内に 226 個存在したことがわかる。226 番目の化合物は、NTG の類似度が 1.0 にも関わらず、元構造の類似度は 0.184 と低かった。これは、この化合物が、環構造を持たない長い側鎖が NTG に付随した構造を持つからである。元構造の類似度は、同型の NTG を持つ 1 番目から 226 番目までの化合物では単調減少しているが、異なる NTG を持つ 227 番目で跳ね上がっている。これ以降も、NTG の

類似度は、順位が下がるにつれて単調減少しており、同型の NTG を持つ場合には、元構造の類似度が単調減少していることがわかる。以上のことから、構造類似性の評価において、NTG の類似度が優先的に評価されていることが確認された。

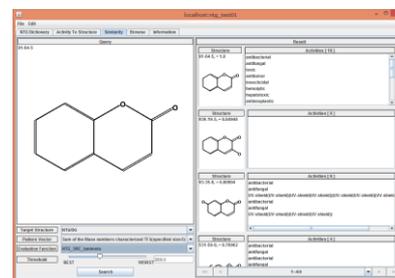


図4 クマリンの構造を入力として構造類似度検索を行った例

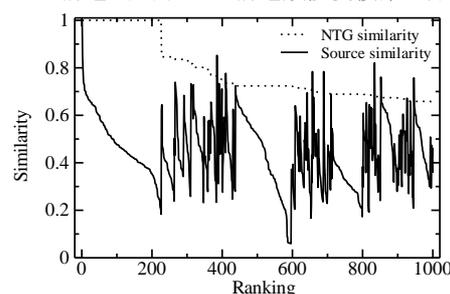


図5 クマリンの構造を入力とした場合の、骨格類似性を優先した構造類似性評価の結果(上位 1000 件)。元構造間、NTG 間の類似性と順位との関係を示す。

5. おわりに

骨格構造を優先した類似性評価手法を、天然有機物の NTG-生理活性知識ベース上に実装し、検索機能を強化した。これによって、天然有機物の構造と生理活性の関係を理解するためのより役立つツールを開発することができた。

参考文献

- [Takahashi 04] Takahashi, Y.: Chemical data mining based on non-terminal vertex graph: In Systems, man and cybernetics, IEEE international conference, pp. 4583-4587 (2004).
- [能登 08] 能登上総, 高橋由雅: 創薬候補構造設計のための NTG 知識ベースの開発, 2008 年度人工知能学会全国大会論文集, p. 212 (2008)
- [Afendi 12] Afendi, F. M. et al.: KNAPsAcK family databases: integrated metabolite-plant species databases for multifaceted plant research, *Plant Cell Physiol.*, Vol. 53, e1 (2012)
- [Nakamura 14] Nakamura, Y. et al.: KNAPsAcK metabolite activity database for retrieving the relationships between metabolites and biological activities, *Plant Cell Physiol.*, Vol. 55, e7 (2014)
- [Takahashi 98] Takahashi, Y. et al.: Structural Similarity Analysis Based on Topological Fragment Spectra, In: R. Carbo and P. Mezey (Eds), *Advances in Molecular Similarity*, Vol. 2, pp. 93-104, JAI Press, Greenwich (1998)
- [Katsuragi 15] Katsuragi, T. et al.: Making an NTG-activity dictionary of bioactive natural organic compounds, The 10th Japan-China Joint Symposium on Drug Discovery and Development, pp. 35-36 (2015)
- [和田 05] 和田雅宏, 高橋由雅: NTG にもとづく骨格構造を優先した類似性検索, 第 33 回構造活性相関シンポジウム講演要旨集, pp. 139-142 (2005)